

## AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

### Listing of claims:

1. (Original) A method for selecting from a plurality of different siRNAs one or more siRNAs for silencing a target gene in an organism, each of said plurality of different siRNAs targeting a different target sequence in a transcript of said target gene, said method comprising

(a) ranking said plurality of different siRNAs according to positional base compositions of a corresponding targeted sequence motifs in said transcript, wherein each said targeted sequence motif comprises at least a portion of the target sequence of the corresponding siRNA and/or a second sequence in a sequence region flanking said target sequence; and

(b) selecting one or more siRNAs from said ranked siRNAs.

2. (Original) The method of claim 1, wherein each said sequence motif comprises said target sequence of said targeting siRNA.

3. (Original) The method of claim 2, wherein said ranking step is carried out by (a1) determining a score for each said different siRNA, wherein said score is calculated using a position-specific score matrix; and (a2) ranking said plurality of different siRNAs according to said score.

4. (Original) The method of claim 3, wherein each said sequence motif is a nucleotide sequence of  $L$  nucleotides,  $L$  being an integer, and wherein said position-specific score matrix is  $\{\log(e_{ij}/p_{ij})\}$ , where  $e_{ij}$  is the weight of nucleotide  $i$  at position  $j$ ,  $p_{ij}$  is the weight of nucleotide  $i$  at position  $j$  in a random sequence, and  $i = G, C, A, U(T)$ ,  $j = 1, \dots, L$ .

5. (Original) The method of claim 3, wherein each said sequence motif is a nucleotide sequence of  $L$  nucleotides,  $L$  being an integer, and wherein said position-specific score matrix is  $\{\log(e_{ij}/p_{ij})\}$ , where  $e_{ij}$  is the weight of nucleotide  $i$  at position  $j$ ,  $p_{ij}$  is the weight of nucleotide  $i$  at position  $j$  in a random sequence, and  $i = G$  or  $C, A, U(T)$ ,  $j = 1, \dots, L$ .

6. (Original) The method of claim 5, wherein said score for each said siRNA is calculated according to equation

$$Score = \sum_{i=1}^L \ln(e_i / p_i)$$

wherein said  $e_i$  and  $p_i$  are respectively weights of the nucleotide at position  $i$  in said sequence motif as determined according to said position-specific score matrix and in a random sequence.

7. (Original) The method of claim 6, wherein each said sequence motif comprises said target sequence of said targeting siRNA and at least one flanking sequence.

8. (Original) The method of claim 7, wherein each said sequence motif comprises said target sequence of said targeting siRNA and a 5' flanking sequence and a 3' flanking sequence.

9. (Original) The method of claim 8, wherein said 5' flanking sequence and said 3' flanking sequence are each a sequence of  $D$  nucleotides,  $D$  being an integer.

10. (Original) The method of claim 9, wherein each said target sequence is a sequence of 19 nucleotides, and each said 5' flanking sequence and 3' flanking sequence are a sequence of 10 nucleotides.

11. (Original) The method of claim 8, wherein each said target sequence is a sequence of 19 nucleotides, and each said 5' flanking sequence and 3' flanking sequence are a sequence of 50 nucleotides.

12-20 (Canceled).

21. (Original) The method of claim 1, further comprising a step of selecting one or more siRNAs based on silencing specificity, said step of selecting based on silencing specificity comprising, (i) for each of said plurality of siRNAs, predicting off-target genes of said siRNA from among a plurality of genes, wherein said off-target genes are genes other than said target gene and are directly silenced by said siRNA; (ii) ranking said plurality of siRNAs according to the number of off-target genes; and (iii) selecting one or more siRNAs for which said number of off-target genes is below a given threshold.

22. (Original) The method of claim 21, wherein said predicting comprises (i1) evaluating sequence of each of said plurality of genes based on a predetermined siRNA sequence match pattern; and (i2) predicting said gene as an off-target gene if said gene comprise a sequence that matches said siRNA based on said sequence match pattern.

23. (Original) The method of claim 22, wherein said step of evaluating comprises identifying an alignment of said siRNA to a sequence in a gene by a low stringency FastA alignment.

24. (Original) The method of claim 23, wherein each said siRNA has  $L$  nucleotides in its duplex region, and wherein said match pattern is represented by a position match position-specific score matrix (pmPSSM), said position match position-specific score matrix consisting of weights of different positions in an siRNA to match transcript sequence positions in an off-target transcript  $\{P_j\}$ , where  $j = 1, \dots, L$ ,  $P_j$  is the weight of a match at position  $j$ .

25. (Original) The method of claim 24, wherein said step (i1) comprises calculating a position match score pmScore according to equation

$$pmScore = \sum_{i=1}^L \ln(E_i / 0.25)$$

where  $E_i = P_i$  if position  $i$  is a match and  $E_i = (1-P_i)/3$  if position  $i$  is a mismatch; and said step (i2) comprises predicting said gene as an off-target gene if said position match score is greater than a given threshold.

26. (Original) The method of claim 25, wherein  $L$  is 19, and wherein said pmPSSM is given by Table I.

27. (Original) The method of claim 26, wherein said plurality of genes comprises all known unique genes of said organism other than said target gene.

28. (Original) The method of claim 10, wherein said position-specific score matrix (PSSM) is obtained by a method comprising

(aa) identifying a plurality of  $N$  siRNAs consisting of siRNAs having 19-nucleotide duplex region and having a silencing efficacy above a chosen threshold;

(bb) identifying for each said siRNA a functional sequence motif, said functional sequence motif comprising a 19-nucleotide target sequence of said siRNA and a 10-nucleotide 5' flanking sequence and a 10-nucleotide 3' flanking sequence;

(cc) calculating a frequency matrix  $\{f_{ij}\}$ , where  $i = G, C, A, U(T); j = 1, 2, \dots, L$ , and where  $f_{ij}$  is the frequency of the  $i$ th nucleotide at the  $j$ th position, based on said siRNAs functional sequence motifs according to equation

$$f_{ij} = \sum_{k=1}^N \delta_{ik}(j),$$

where  $\delta_{ik}(j) = \begin{cases} 1, & \text{if } k = i \\ 0, & \text{if } k \neq i \end{cases}$ , and

(dd) determining said PSSM by calculating  $e_{ij}$  according to equation

$$e_{ij} = \frac{f_{ij}}{N}.$$

29. (Original) The method of claim 28, wherein said plurality of  $N$  siRNAs target a plurality of different genes having different transcript abundances in a cell.

30. (Original) The method of claim 29, wherein said step (b) is carried out by selecting one or more siRNAs having the highest scores.

31. (Original) The method of claim 29, wherein said step (b) is carried out by selecting one or more siRNAs having a score closest to a predetermined value, wherein said predetermined value is the score value corresponding to the maximum median silencing efficacy of a plurality of siRNA sequence motifs.

32. (Original) The method of claim 31, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

33. (Original) The method of claim 29, wherein said step (b) is carried out by selecting one or more siRNAs having a score within a predetermined range, wherein said predetermined range is a score range corresponding to a plurality of siRNAs sequence motifs having a given level of silencing efficacy.

34. (Currently Amended) The method of claim 33, wherein said silencing efficacy is above 50%, ~~75%, or 90%~~ at an siRNA dose of about 100nM.

35. (Original) The method of claim 34, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

36. (Canceled).

37. (Currently Amended) The method of ~~any one of claims 5-11~~ claim 5, wherein said position-specific score matrix (PSSM) is obtained by a method comprising

(aa) initializing said PSSM with random weights;

(bb) selecting randomly a weight  $w_{ij}$  obtained in (aa);

(cc) changing the value of said selected weight to generate a test psPSSM comprising said selected weight having said changed value;

(dd) calculating a score for each of a plurality of siRNAs functional sequence motifs using said test PSSM according to equation

$$Score = \sum_{k=1}^L \ln(w_k / p_k)$$

wherein said  $w_k$  and  $p_k$  are respectively weights of a nucleotide at position  $k$  in said functional sequence motif and in a random sequence;

(ee) calculating correlation of said score and a metric of a characteristic of an siRNA among said plurality of siRNAs functional sequence motifs;

(ff) repeating steps (cc)-(ee) for a plurality of different values of said selected weight in a given range and retain the value that corresponds to the best correlation for said selected weight; and

(gg) repeating steps (bb)-(ff) for a chosen number of times; thereby determining said PSSM.

38. (Original) The method of claim 37, further comprising selecting said plurality of siRNA functional sequence motifs by a method comprising

(i) identifying a plurality of siRNAs consisting of siRNAs having different values in said metric;

(ii) identifying a plurality of siRNA functional sequence motifs each corresponding to an siRNA in said plurality of siRNAs.

39. (Original) The method of claim 38, wherein said characteristic is silencing efficacy.

40. (Original) The method of claim 39, wherein said plurality of *N* siRNAs target a plurality of different genes having different transcript abundances in a cell.

41. (Original) The method of claim 40, wherein said step (b) is carried out by selecting one or more siRNAs having the highest scores.

42. (Original) The method of claim 40, wherein said step (b) is carried out by selecting one or more siRNAs having a score closest to a predetermined value, wherein said predetermined value is the score value corresponding to the maximum median silencing efficacy of a plurality of siRNA sequence motifs.

43. (Original) The method of claim 42, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

44. (Original) The method of claim 40, wherein said step (bb) is carried out by selecting one or more siRNAs having a score within a predetermined range, wherein said predetermined range is a score range corresponding to a plurality of siRNAs sequence motifs having a given level of silencing efficacy.

45. (Currently Amended) The method of claim 44, wherein said silencing efficacy is above 50%, ~~75%, or 90%~~ at an siRNA dose of about 100nM.

46. (Original) The method of claim 45, wherein said plurality of siRNA sequence motifs are sequence motifs in transcript having abundance level of less than about 3-5 copies per cell.

47. (Canceled).

48. (Original) The method of claim 37, wherein said position-specific score matrix (PSSM) comprises  $w_k$ ,  $k = 1, \dots, L$ ,  $w_k$  being a difference in probability of finding nucleotide G or C at sequence position  $k$  between a first type of siRNA and a second type of siRNA, and wherein said score for each said strand is calculated according to equation

$$Score = \sum_{k=1}^L w_k .$$

49. (Original) The method of claim 48, wherein said first type of siRNA consists of one or more siRNAs having silencing efficacy no less than a first threshold and said second type of siRNA consists of one or more siRNAs having silencing efficacy less than a second threshold.

50. (Original) The method of claim 49, wherein said difference in probability is described by a sum of Gaussian curves, each of said Gaussian curves representing the difference in probability of finding a G or C at a different sequence position .

51. (Original) The method of claim 50, wherein said first and second threshold are both 75% at an siRNA dose of 100nM.

52. (Original) A method for selecting from a plurality of different siRNAs one or more siRNAs for silencing a target gene in an organism, each of said plurality of different siRNAs targeting a different target sequence in a transcript of said target gene, said method comprising

(a) ranking said plurality of different siRNAs according to positional base composition of reverse complement sequences of sense strands of said siRNAs; and

(b) selecting one or more siRNAs from said ranked siRNAs.

53. (Original) The method of claim 52, wherein said ranking step is carried out by (a1) determining a score for each said different siRNA, wherein said score is calculated using a position-specific score matrix; and (a2) ranking said plurality of different siRNAs according to said score.

54. (Original) The method of claim 53, wherein said siRNA has a nucleotide sequence of  $L$  nucleotides in its duplex region,  $L$  being an integer, wherein said position-specific score matrix comprises  $w_k$ ,  $k = 1, \dots, L$ ,  $w_k$  being a difference in probability of finding nucleotide  $G$  or  $C$  at sequence position  $k$  between reverse complement of sense strand of a first type of siRNA and reverse complement of sense strand of a second type of siRNA, and wherein said score for each said reverse complement is calculated according to equation

$$Score = \sum_{k=1}^L w_k .$$

55. (Original) The method of claim 54, wherein said first type of siRNA consists of one or more siRNAs having silencing efficacy no less than a first threshold and said second type of siRNA consists of one or more siRNAs having silencing efficacy less than a second threshold.

56. (Original) The method of claim 55, wherein said difference in probability is described by a sum of Gaussian curves, each of said Gaussian curves representing the difference in probability of finding a  $G$  or  $C$  at a different sequence position .

57. (Original) The method of claim 56, wherein said first and second threshold are both 75% at an siRNA dose of 100nM.

58. (Original) A method for selecting from a plurality of different siRNAs one or more siRNAs for silencing a target gene in an organism, each of said plurality of different siRNAs targeting a different target sequence in a transcript of said target gene, said method comprising, (i) for each of said plurality of different siRNAs, predicting off-target genes of said siRNA from among a plurality of genes, wherein said off-target genes are genes other than said target gene and are directly silenced by said siRNA; (ii) ranking said plurality of different siRNAs according to the number of off-target genes; and (iii) selecting one or more siRNAs for which said number of off-target genes is below a given threshold.

59-66 (Canceled).

67. (Original) A method for determining a base composition position-specific score matrix (bsPSSM)  $\{\log(e_{ij}/p_{ij})\}$  for representing base composition patterns of siRNA functional sequence motifs of  $L$  nucleotides in transcripts, wherein  $i = G, C, A, U(T)$  and  $j = 1, 2, \dots, L$ , and wherein each said siRNA functional sequence motif comprises at least a



portion of the target sequence of the corresponding targeting siRNA and/or a sequence in a sequence region flanking said target sequence, said method comprising

(a) identifying a plurality of  $N$  different siRNAs consisting of siRNAs having a silencing efficacy above a chosen threshold;

(b) identifying a plurality of  $N$  corresponding siRNA functional sequence motifs, one for each said different siRNA;

(c) calculating a frequency matrix  $\{f_{ij}\}$ , where  $i = G, C, A, U(T)$ ;  $j = 1, 2, \dots, L$ , and where  $f_{ij}$  is the frequency of the  $i$ th nucleotide at the  $j$ th position, based on said plurality of  $N$  siRNAs functional sequence motifs according to equation

$$f_{ij} = \sum_{k=1}^N \delta_{ik}(j),$$

where  $\delta_{ik}(j) = \begin{cases} 1, & \text{if } k = i \\ 0, & \text{if } k \neq i \end{cases}$ , and

(d) determining said psPSSM by calculating  $e_{ij}$  according to equation

$$e_{ij} = \frac{f_{ij}}{N}.$$

68. (Original) The method of claim 67, wherein each said siRNA functional motif comprises the target sequence of the corresponding targeting siRNA and one or both flanking sequences of said target sequence.

69. (Original) The method of claim 68, wherein each said siRNA has  $M$  nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of  $M$  nucleotides, a 5' flanking sequence of  $D_1$  nucleotides and a 3' flanking sequence of  $D_2$  nucleotides.

70. (Original) The method of claim 69, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides, a 5' flanking sequence of 10 nucleotides and a 3' flanking sequence of 10 nucleotides.

71. (Original) The method of claim 69, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides, a 5' flanking sequence of 50 nucleotides and a 3' flanking sequence of 50 nucleotides.

72. (Original) The method of claim 67, wherein said plurality of  $N$  siRNAs each targets a gene whose transcript abundance is within a given range in a cell.

73. (Currently Amended) The method of claim 72, wherein said range is at least about 5, ~~10, or 100~~ transcripts per cell.

74. (Original) The method of claim 72, wherein said range is less than about 3-5 transcripts per cell.

75. (Currently Amended) The method of ~~any one of claims 67-74~~ claim 67, wherein said threshold is 50%, ~~75%, or 90%~~ at an siRNA dose of about 100nM.

76. (Canceled).

77. (Original) A method for determining a base composition position-specific score matrix (bsPSSM)  $\{w_{ij}\}$  for representing a base composition pattern representing a plurality of different siRNA functional sequence motifs of  $L$  nucleotides, wherein  $i = G, C, A, U(T)$  and  $j = 1, 2, \dots, L$ , and wherein each said siRNA functional sequence motif comprises at least a portion of the target sequence of the corresponding targeting siRNA and/or a sequence in a sequence region flanking said siRNA target sequence, said method comprising

(a) initializing said bsPSSM with random weights;

(b) selecting randomly a weight  $w_{ij}$  obtained in (a);

(c) changing the value of said selected weight to generate a test psPSSM comprising said selected weight having said changed value;

(d) calculating a score for each of said plurality of siRNAs functional sequence motifs using said test psPSSM according to equation

$$Score = \sum_{k=1}^L \ln(w_k / p_k)$$

wherein said  $w_k$  and  $p_k$  are respectively weights of a nucleotide at position  $k$  in said functional sequence motif and in a random sequence;

(e) calculating correlation of said score and a metric characterizing an siRNA among said plurality of siRNAs functional sequence motifs;

(f) repeating steps (c)-(e) for a plurality of different values of said selected weight in a given range and retain the value that corresponds to the best correlation for said selected weight; and

(g) repeating steps (b)-(f) for a chosen number of times; thereby determining said psPSSM.

78. (Original) A method for determining a base composition position-specific score matrix (bsPSSM)  $\{w_{ij}\}$  for representing a base composition pattern representing a plurality of different siRNA functional sequence motifs of  $L$  nucleotides, wherein  $i = G/C, A, U(T)$  and  $j = 1, 2, \dots, L$ , and wherein each said siRNA functional sequence motif comprises a least a portion of the target sequence of the corresponding siRNA and/or a sequence in a sequence region flanking said siRNA target sequence, said method comprising

(a) initializing said bsPSSM with random weights;

(b) randomly selecting a weight  $w_{ij}$  obtained in (a);

(c) changing the value of said selected weight to generate a test psPSSM comprising said selected weight having said changed value;

(d) calculating a score for each of said plurality of siRNA functional sequence motifs using said test psPSSM according to equation

$$Score = \sum_{j=1}^L \ln(w_k / p_k)$$

wherein said  $w_k$  and  $p_k$  are respectively weights of a nucleotide at position  $k$  in said functional sequence motif and in a random sequence;

(e) calculating a correlation of said score and a metric of a characteristic of an siRNA among said plurality of siRNAs functional sequence motifs;

(f) repeating steps (c)-(e) for a plurality of different values of said selected weight in a given range and retain the value that corresponds to the best correlation for said selected weight; and

(g) repeating steps (b)-(f) for a chosen number of times; thereby determining said psPSSM.

79. (Currently Amended) The method of claim 77 [[or 78]], wherein each said siRNA functional motif comprises the target sequence of the corresponding targeting siRNA and one or both flanking sequences of said target sequence.

80. (Original) The method of claim 79, further comprising selecting said plurality of siRNA functional sequence motifs by a method comprising

(i) identifying a plurality of siRNAs consisting of siRNAs having different values in said metric;

(ii) identifying a plurality of siRNA functional sequence motifs each corresponding to an siRNA in said plurality of siRNAs.

81. (Original) The method of claim 79, wherein each said siRNA has  $M$  nucleotides in its duplex region, and wherein each said siRNA functional sequence motif consists of an siRNA target sequence of  $M$  nucleotides and a  $D_1$  nucleotide flanking sequence upstream and a  $D_1$  nucleotide flanking sequence downstream.

82. (Original) The method of claim 81, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides and a 10 nucleotide flanking sequence upstream and a 10 nucleotide flanking sequence downstream.

83. (Original) The method of claim 82, wherein each said siRNA has 19 nucleotides in its duplex region, and wherein said siRNA functional sequence motif consists of an siRNA target sequence of 19 nucleotides and a 50 nucleotide flanking sequence upstream and a 50 nucleotide flanking sequence downstream.

84. (Original) The method of claim 82, wherein said metric is silencing efficacy.

85. (Original) The method of claim 84, wherein said plurality of siRNAs consisting of siRNAs targeting genes whose transcript abundance is in a given range in a cell.

86. (Currently Amended) The method of claim 85, wherein said range is at least about 5, ~~10, or 100~~ transcripts per cell.

87. (Original) The method of claim 85, wherein said range is less than about 3-5 transcripts per cell.

88. (Currently Amended) The method of ~~any one of claims 77-78~~ claim 77, wherein said threshold is 50%, ~~75%, or 90%~~ at an siRNA dose of about 100nM.

89. (Original) The method of claim 84, further comprising evaluating said psPSSM using an ROC (receiver operating characteristic) curve of the sensitivity of said psPSSM vs. the non-specificity of said psPSSM curve, said sensitivity of said PSSM being the proportion of true positives detected using said psPSSM as a fraction of total true positives, and said non-specificity of said PSSM being the proportion of false positives detected using said psPSSM as a fraction of total false positives.

90. (Currently Amended) The method of claim 84, wherein said plurality of siRNA functional sequence motifs consists of at least 50, ~~at least 100, or at least 200~~ different siRNAs functional sequence motifs.

91. (Original) The method of claim 84, further comprising testing said psPSSM using another plurality of siRNA functional sequence motifs.

92. (Original) A method for determining a position match position-specific score matrix (pmPSSM)  $\{E_i\}$  for representing position match pattern of an siRNA of  $L$  nucleotides with its target sequence in a transcript, wherein  $E_i$  is a score of a match at position  $i$ ,  $i = 1, 2, \dots, L$ , said method comprising

(a) identifying a plurality of  $N$  siRNA off-target sequences, wherein each said off-target sequence is a sequence on which said siRNA exhibits silencing activity;

(b) calculating a position match weight matrix  $\{P_i\}$ , where  $i = 1, 2, \dots, L$ , based on said plurality of  $N$  siRNAs off-target sequences according to equation

$$P_i = \frac{1}{N} \sum_{k=1}^N \delta_k(j),$$

where  $\delta_k(j)$  is 1 if  $k$  is a match, and is 0 if  $k$  is a mismatch; and

(c) determining said psPSSM by calculating  $E_i$  such that  $E_i = P_i$  if position  $i$  is a match and  $E_i = (1-P_i)/3$  if position  $i$  is a mismatch.

93-94 (Canceled).

95. (Original) A method for evaluating the relative activity of the two strands of an siRNA in off-target gene silencing, comprising comparing position specific base composition of the sense strand of said siRNA and position specific base composition of the antisense strand of said siRNA or reverse complement strand of said sense strand of said siRNA, wherein said antisense strand is the guiding strand for targeting the intended target sequence.

96-113 (Canceled).